

# NOTES 19: CLT-BASED INFERENCE FOR PROPORTIONS

Stat 120 | Fall 2025

Prof Amanda Luby

---

**CLT:** The Central Limit Theorem (CLT) tells us that if the sample size is big enough and the sample is random,

$$\bar{X} \sim N(\text{_____}, \text{_____})$$

$$\hat{p} \sim N(\text{_____}, \text{_____})$$

The general form for a **confidence interval** is:

**Example:** Finding  $z^*$

- 95% confidence interval
- 68% confidence interval
- 99% confidence interval
- 90% confidence interval

## 0.1 How big is big enough?

**Example 1:**  $\hat{p} = .5$

**Example 1:**  $\hat{p} = .05$

### Rule of Thumb for Proportions:

- Expected count in each category (Yes/No) should be  $> \text{_____}$
- $np > \text{_____}$  and  $n(1 - p) > \text{_____}$

## 0.2 How do we find the SE?

### Standard Error for Proportions:

**Idea:** As  $n$  gets bigger, the SE gets \_\_\_\_\_.

If  $\hat{p}$  is close to .5, the SE is \_\_\_\_\_ than if  $\hat{p}$  is close to 0 or 1

**Example:** ESP Example Again

$n = 14$

```
p_hat = 3/14
n = 14
SE_p = sqrt((p_hat*(1-p_hat))/n)
z_score = (p_hat - .2)/SE_p
p_val = pnorm(z_score, lower.tail = FALSE)
p_val
```

[1] 0.4482

$n=1400$ :

$n=14000$ :

**Big Picture** Picture

## 1 In-class Group Questions

1. APM Research Lab ran a [survey of likely Minnesota voters](#) between Sept 16-18 of 2024. They reported 48.4% in support of the Harris/Walz ticket, and 43.3% in support of the Trump/Vance ticket.

Here's an excerpt from their methodology report:

*The margin for error, according to standards customarily used by statisticians, is no more than ± 3.5 percentage points. This means that there is a 95% probability that the “true” figure would fall within that range if all voters were surveyed.*

- (a) Show how the margin of error was computed
  - (b) Is it OK to use the normal distribution as a model for the sampling distribution of a proportion in this case?
- 
2. In a random sample of 100 moviegoers in January 2013, 64 of them said they are more likely to wait and watch a new movie in the comfort of their own home.
    - (a) Is it ok to use the normal distribution as a model for the sampling distribution of the proportion of moviegoers preferring to wait and watch at home?
    - (b) Suppose I want to know if more than 50% of moviegoers in the population prefer to wait and watch at home. Write down the corresponding null and alternative hypotheses.
    - (c) For the hypothesis test in part (b), find the standardized test statistic (i.e., the Z-score).

- (d) Calculate the p-value and state a conclusion in context
- (e) The table below gives some of the percentiles of the  $N(0,1)$  distribution. If you were given this table (for example on an exam, where you do not have access to R), what could you say about the value of the p-value?
- | percentage | percentile ( $qnorm(percentage)$ ) |
|------------|------------------------------------|
| 90%        | 1.3                                |
| 95%        | 1.6                                |
| 97.5%      | 2.0                                |
| 99%        | 2.3                                |
| 99.5%      | 2.6                                |
- (f) Compute and interpret a 90% confidence interval for the population proportion of moviegoers who believe they are more likely to watch a new movie from home. (Use the information from the previous table to find the critical value needed.)
- (g) How big of a sample size would you need if you wanted the margin of error for a 90% confidence interval to be no more than 2%? To answer this question, assume that the true proportion is  $p = 0.5$ , which would result in the biggest margin of error for a given sample size.